Intelligent Emotional Interface for Personal Robot and Its Application to a Humanoid Robot, AMIET

Yong-Ho Seo, Il-Woong Jeong, Hye-Won Jung, Hyun S. Yang

Department of Electrical Engineering and Computer Science, KAIST, Daejon, Korea
(Tel : +82-42-869-3527; E-mail: {hsyang, yhseo, woong, hwjung}@paradise.kaist.ac.kr)

Abstract: In the near future, robots will be used for the personal use. To provide useful services to humans, it will be necessary for robots to understand human intentions. Consequently, the development of emotional interfaces for robots is an important expansion of human-robot interactions. We designed and developed an intelligent emotional interface for the robot, and applied the interfaces to our humanoid robot, AMIET. Subsequent human-robot interaction demonstrated that our intelligent emotional interface is very intuitive and friendly

Keywords: Human-Robot Interaction, Affective Computing, Humanoid Robot, Multi-modal Interfaces

1. INTRODUCTION

Researchers in robotics have been exploring various interaction methods between humans and robots. Some remarkable studies on robotics and agents have focused on emotion-based interaction with humans. For example, a social robot named Kismet conveys intentionality through facial expressions and engages in infant-like interaction with a human counterpart [1]. In addition, AIBO, an entertainment robot, behaves like a friendly and life-like dog [2]. Shibata's robotic cat was developed to investigate the emotional behavior of physical interaction between a cat and a human [3].

Due to the limitations of current recognition technologies, most of the previous research in robotics has focused on the development of passive robots that interact mainly by responding to their users. Our research concerns a new approach to build a sociable robot that communicates with humans by leading the interaction. Our goal is to enable high-level interactions between humans and robots, such as emotional conversation.

We developed an intelligent emotional interface system for a humanoid robot name AMIET, allowing AMIET to achieve high-level interactions by using proper behavior in response to user's interaction. To make it possible for the robot to lead during interactions, we considered the following objectives. First, the robot must have a social personality to induce interaction and must be able to approach people to start interactions. Second, to enable robust emotional interactions, the robot has to change its behavior according to its perception of a user's emotion. Third, the robot must be able to store memories of previous interactions with people, to direct subsequent interactions more naturally based on previous associations of events and emotions. Fourth, the robot should be able to continue estimating a user's response based on the current context. Moreover, the robot must be able to lead interactions with multimodality.

This paper is organized as follows. In section 2, we present an overview of the designed system. In section 3, we introduce our humanoid robot AMIET. In section 4 to 8, we present the each of the sub-system implementations. In section 9, we present our experimental results and conclusions.

2. SYSTEM OVERVIEW

2.1 Design Concept

The intelligent emotional interface was designed to realize affective interaction with multimodality; Multimodality comprises the multimodal interaction channels through speech, facial expressions, voice and gestures. Our main goal was to make this system proficient at understanding and transferring emotional expressions to a human partner and giving the system a variety of social skills to foster a robot's communicative behavior with humans.

The main tasks of the intelligent emotional interface system are summarized as follows.

- The system determines a behavior appropriate for its current human partner's emotions and behaves autonomously.
- The system recognizes human emotions and expresses its own emotions through multimodal channels
- The system synthesizes its own emotion as an artificial life.

2.2 System Framework

For an intelligent service robot, we designed an emotional interaction system framework to include five subsystems as shown in Fig. 1. Our framework is based on a creature kernel framework for synthetic characters [4]. A similar creature kernel framework was also applied to the software architecture of Kismet [1].

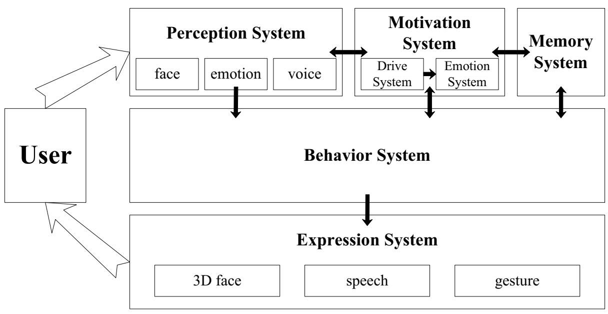

Fig. 1 System Flow

The arrows in above Fig. 1 represent the information flow and influence among the subsystems. Well-coordinated interaction among these subsystems is required for our robot to behave gracefully in a dynamic world.

The functions of each subsystem are summarized as follows.

The perception system is a system that mainly extracts information from the surrounding environments. In our system, the perception system detects whether a human is in sight of the robot, identifies any human in its presence, determines the human's emotional status, and listens for human speech. The perception system is composed of different sub-systems including a face detection system, face recognition system and an emotion recognition system.

The motivation system is composed of drive and emotion system. Drives are motivators, and include internally produced drives and externally induced desires. The robot has three basic drives for accomplishing its tasks and objectives: (1) the drive to interact with humans, (2) the drive to ingratiate itself to humans, and (3) the drive to maintain its vitality. The emotion synthesis system produces the robot's artificial emotions. Currently, three emotions have been modeled to give the robot synthetic analogues of anger, joy, and sorrow.

The memory system saves the most frequent occurring emotion during the last interaction with a particular user, thus, influencing the robot's emotional status when meeting the same person again. Next, the behavior system performs an action selection i.e., it chooses the most relevant behavior to perform in accordance with a given perception, motivation and memory input.

The expression system expresses the emotion of the robot by using appropriate 3D facial expressions, speech and gestures, based on the result of the behavior system.

3. A HUMANOID ROBOT, AMIET

AMIET is the name of a humanoid robot that we designed and developed. It has human-like torso and two wheels in her lower body for navigation. It has two CCD cameras for its vision system. And it has a speaker and a microphone for expressing and recognizing voice. In its chest, there is a LCD screen for expressing its emotional status using graphical representation. The shape of AMIET is shown Fig. 2.

Fig. 2 The shape of AMIET

AMIET can recognize a user's identity using its perception system. It can make human-like motions with its two arms; each arm has 6 degrees of freedom (DOF), so it can imitate the motion of a human arm. Additionally, AMIET has a waist with 2 DOF to perform rotating and bending motions. Thus, AMIET can perform many human-like acts. We used AMIET to test our intelligent emotional interface.

4. PERCEPTION SYSTEM

The perception system comprises a face detection system, face recognition system, a voice recognition system, and an emotion recognition system. Through its perception system, AMIET can collect information about a human user's identity, the user's emotional state and the user's intent. The process of this subsystem is shown Fig. 3.

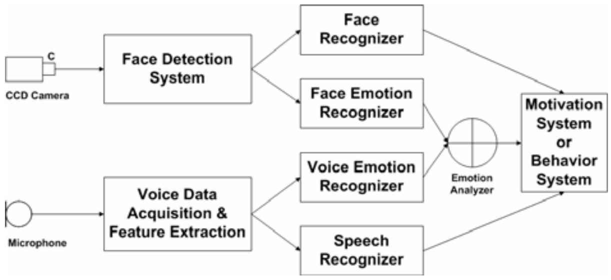

Fig. 3 The Architecture of the Perception System

4.1 Face Detection and Recognition System

The face detection system finds the areas in an image of a human from captured with its CCD camera. For a robust and efficient detection, the face detection system uses a bottom-up, feature-based approach. The system searches the image for a set of facial features, such as color and shape, and groups them into face candidates based on their geometric relationships. Then, the system decides whether the candidate region is actually a face by locating eyes in the eye region of a candidate's face. The detected facial image is used for an input of the face recognition system.

The face recognition system determines the user's identity from the face database. The system was implemented using an unsupervised PCA-based face classifier commonly used in face recognition.

4.2 Emotion Recognition System

The emotion recognition system recognizes the user's emotion using AMIET's detection of the user's facial expression and speech.

To recognize emotions from human facial expressions, a normalized face image captured by the CCD camera is necessary. Facial and edge image features of lips, eyebrows, and forehead are extracted from the normalized image to recognize the user's emotion. AMIET's emotion recognition system is based on Ekman's facial expression features [5].

Each of extracted features is then trained using two neural networks for three emotions – "happy," "sad," and "angry." These emotions are chosen from the basic six emotions defined by Ekman [8] to enhance the recognition results.

To recognize emotions from human speech, we adopted a similar recognition method used in Life-like communication agents called MUSE and MIC [6]. Two kinds of features are used in our emotion recognition system. One is a phonetic feature and the other is a prosodic feature. Each feature vector was trained using a neural network for the three emotions developed – happy, sad and angry. These three emotions were taken as classifiers because these emotions have a higher recognition rate than other emotions such as surprise, fear, or disgust based on the experimental results of previous systems designed to recognize emotions from human speech.

We integrated the training results of the two emotion recognizers mentioned above. The recognizers were integrated using a decision logic for determining current user's emotion from among it three emotions-happy, sad and angry.

The final result vector of the decision logic is calculated as follows

$$R_{final} = (R_{face}W_f + R_{speech}W_s) + R_{final-1} - \delta t \quad (1)$$

Where, R_{face} and R_{speech} are the result vectors of the emotion recognition from the user's facial expression and speech. W_f and W_s are the weights of the two modalities. $R_{final-1}$ is the previous emotion result, as determined by the decision logic and δt is a decay term that eventually restores the emotional status to neutral. The decision logic saves the human emotional status in the final result vector, R_{final}. We constructed the final decision logic to carry out a weighted summation. Accordingly, we optimized the weight valuables, W_f and W_s by experiments. The overall bimodal emotion system yielded approximately 80% for each of five testers. It achieved higher performance results than facial-only and speech-only by resolving some confusion.

5. MOTIVATION SYSTEM

The motivation system establishes AMIET's character by defining its 'needs' and influencing how and when AMIET acts to satisfy them. The character of the robot is to communicate with humans affectively and to ingratiate itself to them.

The motivation system consists of two subsystem: a drive system and an emotion system. Each subsystem serves as a regulatory function for AMIET to maintain its vitality.

5.1 Drive System

In the drive system, three basic drives were defined for the affective communication with humans: a drive to interact with humans, a drive to ingratiate itself with humans, and a drive to maintain its vitality. In the current implementation, the three drives operate as follows

The drive to interact with human motivates the robot to find and approach, and greet a human. Accordingly, if the robot cannot find a human through face detection, its activation intensity increases.

The drive to ingratiate itself to humans prompts the robot to make a human feel better. For example, when the robot interacts with a human, it tries to ingratiate itself while considering the person's emotional state. When the person feels sorrow, the robot attempts to solace the person. If the intensity of the recognized human emotion through the perception system is over a predefined threshold value, its activation intensity increases.

The drive to maintain its vitality is related to the robot's maintenance of its vitality with regard to psychological and physical fatigue. For example, if the robot expresses extreme anger or sadness, it stops interacting with the human. The psychological and physical condition is expressed as an 'energy' variable. If the energy is too low, its activation intensity increases.

5.2 Emotion Synthesis System.

Emotions play an important role in human behavior, communication and interaction. Accordingly, the robot's emotions also play an important part in our system. First, the robot expresses its emotional status using 3D facial expressions, speech and gestures. Second, the synthesized emotion influences the behavior system and the drive system as a control mechanism.

To synthesize AMIET's emotions, we used the emotion model that characterizes emotions using the three dimensions of stance, valence, and arousal [7]. So, it is possible to derive emotions from physical variables.

Our system is under open stance because AMIET is motivated to be openly involved in interaction with humans.

The arousal factor is determined by a human and his or her responses, that is whether AMIET finds a human or not and whether his response is or not. For example, if AMIET cannot find any human, the arousal decreases. A low arousal increases AMIET's sad emotion while a high arousal increases AMIET's happy or angry emotion, by determining whether the human response is positive or not.

The valence factor is determined by human's response, that is how the human responds appropriately to the robot's behavior. For example, when AMIET waits for an answer, yes or no, if human says something unexpected, the valence factor becomes negative. A negative response increases the anger emotion while a positive response increases the happy emotion.

The robot's emotional status is computed by the following equation.

$$E_i(t) = \begin{cases} M_i, & t = 0. \\ A_i(t) + E_i(t-1) + \sum D_i(t) - \delta t, & t \neq 0. \end{cases}$$

($t = 0$, when a new face appeared) (2)

Where E_i is the robot's emotional status, t is time, i = {joy, sorrow, anger}. A_i is the emotional status calculated by the mapping function of [A, V, S] from the current activated behavior. D_i is the emotional status defined by the activation and the intensity of unsatisfied drives in the drive system. M_i is the emotional status of the human recorded in the memory system. Finally, δt is a decay term that eventually restores the emotional status to neutral.

6. MEMORY SYSTEM

Our system stores memories required for a more natural and intelligent communication with a human.

Humans often have feelings for those with whom they have communicated in the past. When a human meets a particular person again, he or she might be influenced by the remembered feelings and other memories of the person. The robot saves the most frequently occurring emotion from the latest interaction with a user. Therefore, the memories about a particular human influence the robot's emotional status when the robot meets the person again.

The robot saves a user's personality traits, such as being active or passive. This information can help the robot to select the proper interaction activity to use with the user. For example, the robot could play quiet music to a passive user while the robot could play rock music to an active user. User's preference saves such as likes and dislikes of several topics in dialog. If the robot mentions such information to the user during their next interaction, the user might feel that the robot is intelligent, making the user more interested in interacting with the robot.

7. BEHAVIOR SYSTEM

The behavior system organizes its goals into a cohesive structure as shown Fig. 4. The structure has three levels and branches that address the three drives described in section 5.

Each branch has multiple levels with three layers. As the system goes into a lower layer, a specific behavior is determined by the affective relationship between the robot and the human.

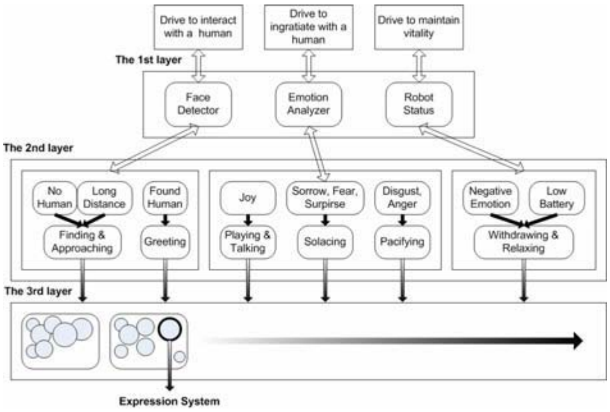

Fig. 4 The behavior system hierarchy

The first layer determines which drive should be selected from the three basic drives. The most important drive was designed to be selected frequently and to effectively activate the proper behavior.

The larger the magnitude of the drive and the more frequently that need must be selected, the grater the contribution that the drive makes to the activation of the behavior. However, the drive to maintain its vitality has the highest priority for activation because this drive is related to the operation power of the robot.

The second layer determines which high-level behavior should be performed, according to the perception and internal information for the selected drive. In the first drive, if the human is far away or absent, a seek and approach behavior is performed; if a human is detected, a greeting behavior is performed. In the second drive, one of the three behaviors is performed, according to the user's emotional state. In the third drive, if the robot is extremely angry or sad, a withdraw and rest behavior is performed.

The third layer consists of some low-level behavior that corresponds to high-level behavior. All low-level behavior is composed of dialog and gestures. This low-level behavior is executed in the expression system. The low-level behaviors of the same behavior group consist of different dialogs, though they have the same behavior goal. Accordingly, a low-level behavior is randomly selected from the behavior group.

8. EXPRESSION SYSTEM

The expression system comprises three sub-systems: a dialog expression system, a 3D facial emotion expression system, and a gesture expression system.

The expression system plays two important functions. The first function is to execute a particular behavior received from the behavior system. An expression behavior consists of a dialog between the robot and a human. Sometimes the robot utilizes interesting gestures to control the dialog's flow and to foster interaction with the human. The second function is to express robot's emotion. The robot expresses its own emotion through facial emotion expressions and gestures.

8.1 Dialog Expression

Dialog is a joint process of communication, the sharing of information between two or more parties. In addition, humans employ a variety of paralinguistic social cues to regulate the flow of dialog. We consider there to be three primary types of dialog: low-level (pre-linguistic), non-verbal and verbal language. Among these types, AMIET can communicate with a human using daily verbal language, along with proper gestures.

However, it is difficult to enable a robot to engage in a natural dialogue with a human because of limitations with the current technology for speech recognition and natural language processing. Thus, we use a predefined dialog flow and topics. In order to make a natural dialog possible within the limits of AMIET's language recognition capabilities, we constructed a dialog as follows: First, AMIET actively leads a dialog by asking a user's intention ahead of him to avoid a situation that AMIET cannot understand. Second, to avoid an unnatural dialog, AMIET answers with its the most frequently used responses when it cannot understand the dialog.

The dialog expressions comprise the most commonly used speeches according to the selected task behavior, as described in section 7. For instance, for the greeting task, AMIET says "Hello" to a user and then asks his or her name. In the talking task, a dialog consists of various common topics, such as hobbies, weather, and movies.

AMIET's speech synthesis utilizes a text-to-speech program developed by Cowon Systems, and this program synthesizes AMIET's voice from the given Korean or English text.

8.2 Facial Expression

The 3D facial expression shows the robot's emotional status synthesized in the motivation system, as described in section 5. These expressions make up for the limitations of the robot's mechanical face which has difficulty in expressing its emotions. These facial emotion expressions were implemented using 3D graphics. Our 3D graphical face is displayed on the LCD screen which located on the robot's chest. The facial expressions in our 3D graphical face are shown as Fig. 5.

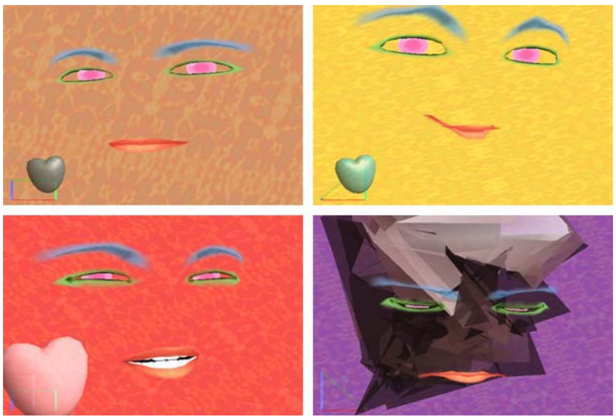

Fig. 5 3D Graphical Facial Expressions

8.3 Gesture Expression

AMIET's gesture were generated to be human-like and friendly. AMIET's gestures are used to express its own emotions and to make interaction with humans more expressive. Therefore, expressions that would best attract the interest of humans were considered, and various interesting gestures were developed for AMIET that would match the robot's dialogs and emotional statuses.

To generate emotional gestures, we used the concept of Laban Movement Analysis [8], which is used for describing body movements. Using Laban Movement Analysis, we extracted some parameters to classify certain characteristics of motions. By adjusting these parameters, we generated some gestures for AMIET with different emotional meanings. Some examples of these gestures are shown in Fig. 6.

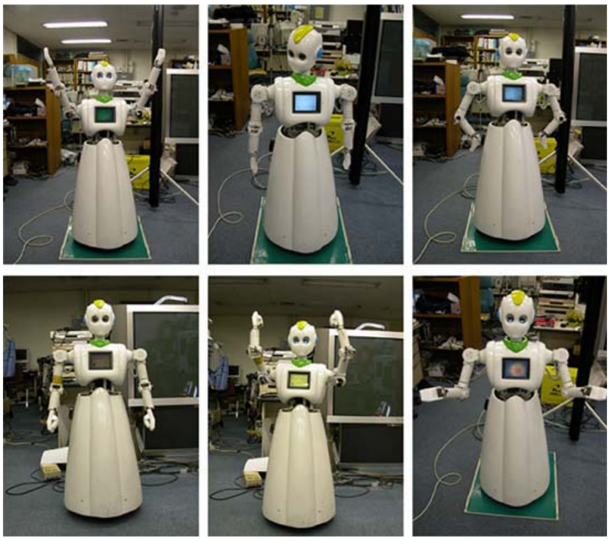

Fig. 6 Gesture Expressions

9. RESULTS & CONCLUSION

We verified the following factors based on the recorded internal parameters and the working status of each sub-system:

- System recognition of human emotional status
- System adjustment of the robot's expressed emotion based on a user's emotion and interaction responses
- System activation of an appropriate drive, based on the activation condition of each drive
- System storage of the robot's previous emotions about different, individual users
- System determination of appropriate behavior, based on the current activated drive and the user's emotional status

We confirmed that each sub-system is worked properly towards satisfying its objectives. Based on the evaluation result, we draw a graph shown as Fig. 7, which shows the flow of the sub-systems about a sample interaction.

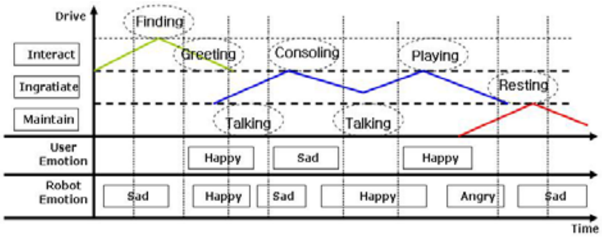

Fig. 7 System Working Flow

Our plans for the future work include enhancement of our robot's emotional expressions. To make the robot more intelligent, it is necessary to increase the number of emotions that the robot can recognize and express. Our current framework is designed to interact with people with only just three emotions – "happy," "sad," and "angry." We will develop more emotions to the robot such as "fear," "disgust," and "surprise."

Fig. 8 AMIET interacts with a human user

REFERENCES

[1] C. Breazeal, *Designing Sociable Robots*, MIT Press, 2002.
[2] Ronald C. Arkin, Masahiro Fujita, Tsuyoshi Takagi and Rika Hasegawa, "An Ethological and Emotional Basis for Human-Robot Interaction", *Robotics and Autonomous Systems*, 42(3-4), March 2003.
[3] T. Shibata, at al., "Emergence of emotional behavior through physical interaction between human and artificial emotional creature", *Int'l Conf. on Robotics and Automation*, pp. 2868-2873, 2000.
[4] SY Yoon, RC Burke, BM Blumberg, GE Schneider, "Interactive Training for Synthetic Characters", *AAAI2000*.
[5] P. Ekman and W. V. Friesen, *Facial Action Coding System Investigator's Guide*, Consulting Psychologists Press, 1978.
[6] Naoko Tosa and Ryohei Nakatsu, "Life-like Communication Agent – Emotion Sensing Character 'MIC' & Feeling Session Character 'MUSE'", *ICMCS*, 1996.
[7] H. Schlossberg, "Three dimensions of emotion", *Psychology Review* 61, 1954.
[8] Toru Nakata, Taketoshi Mori & Tomomasa Sato, "Analysis of Impression of Robot Bodily Expression," *Journal of Robotics and Mechatronics*, Vol.14, No.1, pp.27--36, 2002.
[9] Hye-Won Jung, Yong-Ho Seo, Jin Choi, Hyun S. Yang, "Development of a Sociable Robot to Realize Affective Human-robot Communication with Multimodality", *VSMM 2003*, Canada.